The Gateway Business Plan

An NCDDC White Paper – November 2002

GATEWAY BUSINESS PLAN

Data that describes the interaction between land and sea along America's coasts is extremely diverse. Coastal data is collected, processed, and maintained by a variety of organizations that are geographically distributed throughout the United States. These data sets exist in many different formats, have different levels of quality assurance and vary substantially in accessibility. Because the cost and complexity of storing and serving data has dropped dramatically in the past decade, most coastal science and management organizations have elected to keep their data and serve it directly to their constituents. The diverse and distributed nature of coastal data, together with the advent of internet and computer technologies, have made the concept of a central coastal data repository impractical. Yet, the need for access to local data for regional, national, and global studies has increased as we have recognized that most coastal problems extend far beyond traditional natural and jurisdictional boundaries.

To meet these challenges and take advantage of new information technologies, the National Coastal Data Development Center (NCDDC) employs a "gateway" concept to provide access to coastal data that can be transmitted over the internet. A "gateway" is a software link between the data provider's server and the NCDDC coastal data "portal." Data consumers, employing the NCDDC web portal and gateways can quickly bring together many types of data, from many sources, to answer specific questions. Gateways can be established between NCDDC and federal, state, and local government agencies; academic institutions; non-profit organizations, and private companies. Together with standardized documentation, or metadata, gateways provide a virtual database of coastal data sets. Ready access to the coastal data network will support decision-making policy, analysis of long-term change, post-event assessment, monitoring, and prediction of coastal phenomena.
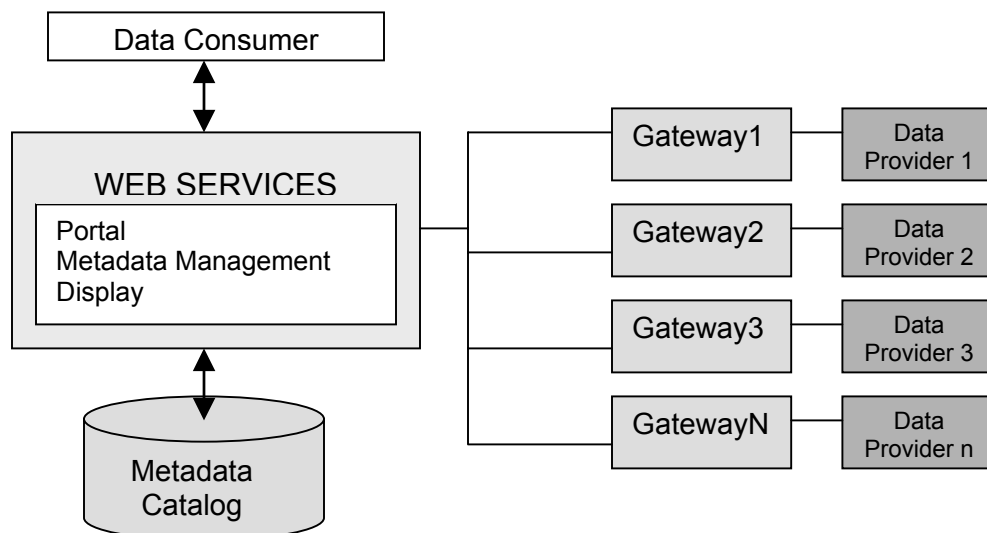
NCDDC works closely with coastal data providers to ensure that all data delivered through the gateways are documented according to Federal Geographic Data Committee (FGDC) and other applicable metadata standards. NCDDC maintains a metadata catalog to help users of the portal to identify relevant data, evaluate its validity and usefulness, and access it.

Four regional liaison officers help data providers and data customers make the best use of NCDDC's information technology resources. They develop relationships with data providers, render assistance to data customers, and provide initial guidance for gateway development.

- Mr. Patrick Caldwell, NCDDC Hawaii Liaison, Dept. of Oceanography, University of Hawaii at Manoa, Honolulu, Hawaii 96822, Phone: (808) 956-4105
- Mr. Michael Crane, NCDDC East Coast Liaison, 4301 Rickenbacker CSWY, Miami FL 33149 Fax: 305 361 4449; Phone: (305) 361 4305
- Mr. Jeff Jenner, NCDDC Gulf of Mexico Liaison, Bldg. 1100, Suite 101, Stennis Space Center, MS 39522. Phone: (228) 688-1579
- Mr. Sidney Stillwaugh, NCDDC West Coast Liaison, 7600 Sand Point Way NE Seattle, WA 98115. Phone: (206) 526-6263

The success of the portal-gateway concept depends on the number of gateways built and metadata records maintained. Initially, NCDDC software engineers will create gateway software and customize it according the unique needs of each data provider. A long-term goal of NCDDC is to develop gateway experts at key institutions throughout the country who will continue to create new gateways in their respective regions.

What is a gateway?

A "gateway" is a software adapter that allows NCDDC to access data and information located on servers outside of NCDDC's computer architecture. Gateways are customized to adapt a particular data source to the NCDDC Data Exchange Infrastructure (DEI) middleware architecture. Through a gateway, NCDDC can deliver data to a client in its native format or, when used in conjunction with the NCDDC Spatial Data Model (SDM), can deliver it in other formats supported by the SDM. Gateways can deliver subsets of data according to criteria specified by the requestor (keywords, time period, geographic area, etc.). Gateways use CORBA (Common Object Request Broker Architecture) as their method of communication. They are "lightweight" (requiring minimal installation of supporting software) and require minimal administration.

Where is a gateway located?

Gateway software is normally located either within the NCDDC computer architecture or on the data provider's computer. Ideally, a gateway is installed as close as possible to the data source-- preferably on the same machine as the data.

If it is appropriate for the gateway software to be installed on the data provider's server, it will normally consist of a C++ executable file (for communications) and a Java executable file (for data adaptation), both written by NCDDC. Supporting software, such as the Java Virtual

Machine, may be required to run the gateway. The source code for each NCDDC executable file is available to the data provider.

The nature of the access to the data and the capabilities of the data provider's server will normally determine the location of the software. If the data provider's server cannot subset the data, it makes sense to locate a gateway with (or near) the data to conserve network bandwidth. For small data sets that may be stored at locations with limited computer and personnel resources, it may make sense to put the gateway software where there are more resources available, such as at NCDDC.

Who can build a gateway?

A gateway programmer should have a working knowledge of the data provider's data structure and server. He or she must be proficient in C++, CORBA, and preferably Java. The fundamental task for a gateway programmer is to write the code to convert the data from its native format to a format that can be sent to NCDDC through the gateway (the externalized SDM representation).

The Gateway Development Kit (GDK) is open source software that can be downloaded from the NCDDC website. This software must then be customized to interface with the unique architecture of the provider's data structure and server architecture. NCDDC will provide a service to build gateways within the center and provide technical expertise to other programmers in the community who want to build their own gateways to data their organizations hold.

What is the Spatial Data Model (SDM) and what is it used for?

The Spatial Data Model is a set of software components for data representation and manipulation. The SDM brings data from different formats together into a common representation that allows data manipulations such as projection and format translations, aggregation, and subsetting. NCDDC will continue to develop application services using the SDM to provide additional manipulation capabilities as needed by customers and NCDDC project managers.

What else is required?

It is equally important to provide for the discovery of the data that is accessible through the gateways. NCDDC provides a searchable metadata catalog that describes the data and provides the connection to the access gateway. The NCDDC catalog will have metadata records for all data accessible through the gateways. Ideally, NCDDC will periodically "harvest" or copy the metadata from the data provider's server to keep the NCDDC version current. If the metadata does not exist or needs to be updated, NCDDC will work with the data provider to produce the required metadata. NCDDC will provide metadata training, tools, and a remote metadata management capability to allow data providers to manage their metadata within the NCDDC system.

Why would I want to generate metadata and build a gateway?

Metadata and gateways make the data holdings available to a broader user community without giving up individual control of data access, quality assurance and distribution. Also, because the gateway can range from direct database access to an HTML query through an existing web site, making data available through the NCDDC portal does not require a data provider to abandon his or her existing web site or other direct data delivery option.

What about my firewall?

Data integrity is a paramount concern for NCDDC as well as the data providers. Along with the great diversity of organizations producing coastal data comes a great diversity of firewall implementations, security policies, and sensitivities to data access by people outside the organizations. There is no single set of firewall requirements or specifications for NCDDC gateways. NCDDC gateway developers work with data providers on a case-by-case basis to work through individual firewall issues.

How does NCDDC decide on which gateways to build?

NCDDC currently supports the development of metadata and gateways for eight pilot or "partnership" projects: Harmful Algal BloomS Observing System (HABSOS); Coastal Risk Atlas (CRA); Fish Habitat; Hypoxia; Integrated Ocean Observing Systems (IOOS); Marine Invasive Species (MIS), Coral Reef Information System (CORIS); and Homeland Security. Data sets and the associated metadata considered fundamental to NOAA's mission receive the highest priority when considering the development of gateways. Accessibility and condition (degree of documentation, quality assurance, etc.) of the data also play a role in establishing gateway development priority. The type of database used to store the data and the difficulty in recovering and making the data available are important factors as well. "Foundation level" data (information cross-cutting various oceanographic and atmospheric programs or projects such as bathymetry, hydrography, coastline, imagery, bottom types, wetlands, watershed information, and coastal infrastructure) normally receives high priority. Balancing gateway development among federal, state, local, academic, non-profit, and private organizations is also important.

NCDDC is committed to building five new gateways per quarter in Fiscal Year 03. The NCDDC staff expects this number to increase with experience in FY04 and beyond. For further technical information on the NCDDC Gateway Business Plan contact the your regional liaison officer or the NCDDC leads:
- Russ Beard, Chief Science Officer (228) 688-3026
- Susan Starke, IT Operations Chief (228) 688-2788